

第4回データサイエンスコンペティション FAQ

Q01：1チームの参加学生数は5名まで可能ですか？

A01：説明会資料の誤りで、1チームの参加人数は2～5名です。

(ただし卒業論文または修士論文に直接関連する分析に限り1名で参加可)

説明会資料を7/8に修正しました。

Q02：申し込みには本学専任教員の下承が必要とありますが、指導教員はどのように決めるのでしょうか？参加者が各自で学科教員に依頼する、あるいはデータ科学センターの教員が担当するのでしょうか？

A02：今回のデータサイエンスコンペティションは、企業から提供いただくデータを利用したコンペティションとなります。企業の提供データの利用にあたっては、データ利用の規約上、教員の管理下にあるPCを用いて分析を行う必要があります。データ科学センターの教員で受け入れることは困難であるため、上記をご理解いただいた上で身近な学部・学科の先生の中で、お引き受けいただける先生（本学専任教員）をご自身で探してください。

Q03：発表会はオンライン形式でしょうか、対面形式でしょうか。

A03：11/26の発表会ですが、コロナの状況等を見ながら対面・オンラインのどちらの形式で実施するかを決定する予定です。実施形式は、10月以降に決定し、参加者に改めてご連絡を差し上げます。

→今回は**オンラインでの開催**として開催いたします（11/1）

Q04：同一人物が2チームに所属することは可能でしょうか

A04：チーム対応のコンペティションとなりますので、お一人につき、1チームのみに参加する形でお願いいたします。

Q05：当日参加できないメンバーがでる可能性があるのですが、その場合でも参加可能でしょうか

A05：チームの他のメンバーで発表可能であれば、問題ありません。

Q06：今回のコンペでは、データ科学センターからクラウドデータ分析環境は提供されますか？

A06：今回のデータサイエンスコンペティションにおいては、データ科学センターでは、コ

ンペ用のクラウド環境の提供は行いません。

また、学生が独自に構築したクラウド環境での分析は、セキュリティや情報漏洩のリスク管理を行えていることを確認することができませんので、今回のコンペでは利用不可とします。

企業の提供データの利用にあたっては、データ利用の規約上、教員の管理下にある PC を用いて分析を行う必要がありますので、ご了承ください。

なお、大学でもウィルス対策ソフトを含めたソフトウェア貸出を行っておりますので、あわせてご確認ください。

早稲田大学 IT サービスナビ

ソフトウェア貸出

<https://www.waseda.jp/navi/rental/soft/index.html>

Q07：今回のコンペで提供いただくみずほ銀行のデータについて、データサイエンスコンペ以外の目的（例えば自身の卒業論文等）で提供データを利用することは可能でしょうか？

A07：提供データは本年度の卒論・修論に利用できます。

卒論・修論に利用する際に必要な手続きの詳細につきましては、データサイエンスコンペで提供されるみずほ銀行のデータを本年度の卒論・修論に利用する際の申請手続き及び注意点につきましては以下をご参照ください。

<申請手続>

卒論・修論に関して学外向けの刊行や外部聴講がある場合は、データ科学センターを通してみずほ銀行に事前の承諾を得る必要があります。

卒論・修論の提出日（例えば、学部・研究科の定める提出〆切日）の2週間前までに、以下のものを指導教員の杉澤先生よりメールにて提出いただくようお願いいたします(注1)。

- ・学外発表申請書（本メールに添付）
- ・本論，発表原稿等

なお、学部・研究科への提出は、みずほ銀行の承諾を得た後に行っていただくようお願いいたします。

学外向けの刊行や外部聴講がない場合、申請手続きは必要ありません。

送付先

データ科学センター WIRP 担当 安田，野村

cds-wirp@list.waseda.jp

(注 1)「個別のデータの利用にかかる遵守事項」

<<https://www.waseda.jp/inst/cds/assets/uploads/2022/07/b2e2bd6cd160ac0362e6c4d47ab6e79a.pdf>> の内容に則ります。

<注意点>

誓約書

<<https://www.waseda.jp/inst/cds/assets/uploads/2022/07/97dac16bd83e19d5be964b20d9a2fbd5-6.pdf>>の内容を遵守して作成ください。

Q08:クレジットカード決済統計データに関して

データ説明資料において銀行データ③のデータの中に「募金」という項目がありますが、その概要を教えてくださいませんか。この項目の中には、いわゆる「募金」(ユニセフ、赤い羽募金など)以外に、「ふるさと納税」なども含まれていますでしょうか？

もし、ふるさと納税は別分類となる場合は、具体的にどの項目に分類されるのでしょうか？

A08:お書きいただきました通り「ユニセフ、赤い羽募金」のような「募金」の概念の決済をこの「募金」の項目にて計上しております。分類「その他/募金」には、いわゆる「募金」の他に「ふるさと納税」も含まれています。"

Q09:口座出金統計データ・クレジットカード決済データの平均値と分散の算出方法に関して個人の特定を避けるため今回のデータは地域内での1ヶ月1人あたりの平均額だと認識しておりますが、その算出過程においての質問がございます。平均値等の計算には決済や支払いが1円でもあった人のみのサンプルで計算されているのでしょうか。例えば支払額が0円の人や0円として認識されているのか(パターンA)、支払いがない場合はその人のデータは欠損値(パターンB)となって、平均や分散の計算に参入されないのかという質問です。

例) A市に2名が居住、うち1名は2020年1月にキャッシュレス決済チャージを1000円、もう1名はチャージ額0の場合

→パターンAの平均値:500円

→パターンBの平均値:1000円

A09:お示しいただきましたパターンBの手法にて、各分類に該当する支払いが1回以上発生しているお客さまのみを対象として平均値・分散の算出を行っています。

Q10:クレジットカード決済統計データに関して

データ説明資料資料において銀行データ③において生活(サービス)の中に「テレビ」という項目がありますが、その概要を教えてくださいませんか？

この「テレビ」という項目が示しているのは、1 テレビ機器自体の購入なのか、あるいは 2 NHK 受信料のようなテレビサービスへの対価の支払いなのかを伺いたいです。

A10:お示しいただきました 2 の通り、テレビサービスへの対価の支払いを示した分類となっています。

Q11:クレジットカード決済統計データの中に「会費」という項目がありますが、こちらの概要も教えてください。

1 サブスクリプションはこの項目に入っているのか。

2 NHK 受信料は「会費」に含まれるのか。

A11:①サブスクリプションサービスの種類にもよりますが、原則サブスクリプションは「生活（サービス）/会費」に含まれております。（Amazon プライムなど）

②NHK 受信料は「生活（サービス）/会費」ではなく、分類「生活（サービス）/テレビ」に含まれております。

Q.12:提供データの欄に「エリアにおけるクレジットカードの利用者数から～」と記載がありますが、こちらは、クレジットカード決済加盟>店エリアにおける利用者数なのか、クレジットカード利用者の住所エリアにおける利用者数なのか、どちらになりますでしょうか。データの主体が、決済を受けた加盟店なのか、決済をした利用者なのか分かりかねます。

A12:「クレジットカード利用者の住所エリアにおける利用者数」になります。

Q13:リスク資産データについて、「エリア毎に『平均』と『分散』を算出」とありますが、こちらの平均の分母は、

a. みずほ銀行の口座を開設している人の総数

b. a.の中で、5 種類のリスク資産それぞれについて、その資産を保有している人の数（例えば、a.の中で投資信託を行なっている人の数）

のどちらでしょうか。

また、リスク資産データと同様に、ストック資産データとローン統計データにつきましても、平均の分母の取り方を教えてください。

A13:「b」になります。ストック資産データとローン統計データにつきましても、同様の取り方です。

Q14:各々の月末残高についてエリア毎に「平均」を算出するために、分母は各エリアみずほ銀行の口座を持っているすべての方の数か、みずほ銀行の特定の金融資産（投資信託、保険）を保有している方だけの数か、もしくはエリアの住民数ですか

A14:「みずほ銀行の特定の金融資産（投資信託、保険）を保有している方だけの数」

分母としてはこちらが回答になります。投資信託の平均・分散、であれば、投資信託を保有している方が分母になりますし、保険の平均・分散であれば、保険を保有されている方が分母になります。

Q15:「みずほ銀行では、複数の保険会社の様々な保険商品を取り扱っております。みずほ銀行 取扱保険商品は年金保険・養老保険 終身保険 定期保険 医療保険 がん保険 介護・認知症保険 就業不能保険 住宅用火災保険 海外旅行保険を含まれています。」と書いてあるですが、今回提供するされた保険というリスク資産は具体的にこれらの商品のどれを特定されていますかもしくは全部の保険商品の平均ですか

A15:保険につきましては「生命保険の一時払い商品」のデータとなります。

よって平準払いの保険（毎月支払うもの）は含まれておらず、損害保険も含まれていないため「住宅用火災保険、海外旅行保険を除く、年金保険・養老保険 終身保険 定期保険 医療保険 がん保険 介護・認知症保険 就業不能保険の一時払い商品」となります。

Q16:みずほ銀行が提供されている住宅ローンは「みずほ住宅ローン、みずほ住宅借り換えローン、買い換えローン、自然災害支援ローンなど」と書いてありますが、今回提供されたデータの中、具体的にどういった住宅ローンの平均値が含まれているのでしょうか。特定の住宅ローン数個だけを基に計算されているか、それとも上記のすべての住宅ローンの商品の平均値になりますでしょうか。

A16:「住宅ローン残高」は、「みずほ住宅ローン、みずほ住宅借り換えローン、買い換えローン」(*)になります。「自然災害支援ローン」は正式名称「自然災害時支援特約付き住宅ローン」で、(*)の借入時に付ける保証プランの一つなので、みずほが提供する住宅ローンの上記商品(*)ととらえてください。

Q17:住宅ローンの返済プランも個人の選択によって異なると思いますが（返済額一定／毎年変わる／返済途中で変えられるような様々なプランがみずほ銀行のHPに書いてあります）、今回のデータでは、これらのプランを全部まとめて平均値を計算されるのでしょうか。

A17:返済プランに関わらず、(*)をお借入れの方全量で平均値を算出しています。

Q18:住宅ローン残高のデータ自体は每期毎期に新たにローンを借りる方とこれまでにすでにローンを借りた個人のローンの月末残高という 2 パターンが考えられますが、今回のデータは新規と既存の方の両方が含まれているのでしょうか。

A18:ご認識の通り、「住宅ローン残高」は新規で借入された方、以前から借入されている方まとめた数字になります。

Q19:データを分析するにあたって今頂いている情報よりも昔の情報を提供して頂ければと思います。

具体的には今頂いている ATM の市区町村別の時間による引き出し額、回数、顧客数のデータ (ATM_Var2_KEKKA) について 2012 年から 2019 年のものを利用したいと考えております。

A19:お問い合わせの件ですが、今回のコンペティション開催にあたり、データサイエンスコンペティション事務局とみずほ銀行との間で提供するデータの範囲を定めており、データ提供範囲外のデータを追加で提供することはできません。

また、コンペの規則からも自由に利用できる以外の追加データの利用は認められないこととなっております。

以 上